2022 The 4th Asia Energy and Electrical Engineering Symposium

Energy Saving in 5G Substations Using Edge Caching for Adaptive Immersive Media Streaming

Emmanuel Osei-Mensah
School of Information and Communication Engineering
University of Electronic Science and Technology of China
Chengdu, China
ecoseimensah@gmail.com

Saqr Khalil Saeed Thabet
School of Information and Communication Engineering
University of Electronic Science and Technology of China
Chengdu, China
saqrthabet2004@yahoo.com

Olusola Bamisile
College of Nuclear and Automation Engineering
Chengdu University of Technology
Sichuan P.R., China
Boomfem@outlook.com

Emelia Asiedu-Ayeh
School of Management and Economics
University of Electronic Science and Technology of China
Chengdu, China
easieduayeh@yahoo.com

Victor Kwaku Agbesi
School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu, China
victoragbesivik@gmail.com

Jian Li
School of Mechanical and Electrical Engineering
University of Electronic Science and Technology of China
Chengdu, China
leejian@uestc.edu.cn

2022 4th Asia Energy and Electrical Engineering Symposium (AEEES) | 978-1-6654-7914-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/AEEES54426.2022.9759596

***Abstract*—As the number of wireless network-dependent devices exponentially grows, wireless access networks require capacity expansion to meet the high bandwidth demands of the applications running on these devices. This results in high power consumption, especially for services like traditional video and immersive media streaming which necessitate stringent high transmission rates. In this study, a part-based media edge caching scheme is proposed for energy-efficient wireless access networks. Intelligent-edge-assisted networks store only parts of the media content on edge servers and deliver the remaining parts when active users continually engage the streaming session. This is done by responding to network queries seeking to continue with content delivery or terminate the streaming session. The reward for a timely response is an assurance of uninterrupted streaming sessions. In this way, a streaming session without a user's response is halted to save the transmission energy and bandwidth of the edge servers. We formulate the edge caching problem as a multi-constrained multiple-knapsack problem and attempt to optimally place the content with the help of the proposed algorithm. Extensive simulations reveal that the proposed caching scheme shows improvements in the cache hit ratios and average delays as compared with LFU, FIFO, and LRU.**

Keywords*—energy efficiency, renewable sources, edge caching, green communications.

I. Introduction

Recently, the advent of immersive media technology is gaining a lot of attention in the technology landscape. Immersive media (Virtual, Augmented, and Mixed realities, and 3D content), also known as extended reality, gives people the degrees of freedom to interact with content on another level [1]. Rather than just watching a traditional 2D video, immersive media can create an environment allowing users to stop into that video and interact with things within the realm. The market share of this technology is projected to be worth $180 billion by the end of 2022 [2]. However, streaming immersive media using the existing network infrastructure faces deepening challenges which make its earnest realization a dream for the future. To begin with, a seamless delivery is dependent on reliable and high bandwidth requirements, and high storage capacities to pre-cache them [3]. Additionally, the preprocessing and further transcoding from a high bitrate resolution to a lower one meant to serve most users with fluctuating wireless access are computationally intensive and costly.

The information-communications technologies (ICT) industry contributes up to 3.9 % of all global carbon emissions [4]. The development of the ICT industry has been one of the most progressive technological and societal developments of the last decade. The ICT industry does not only constitute an industry in its own right but also connects virtually to every other business sector, all of which constantly rely on ICT for a range of needs. It is anticipated that the ICT companies may receive continual pressure on network energy efficiency because the energy of the sector is still increasing, with high expectations concerning sustainable development and environmental impact reductions. Future networks are projected to run fully or partially on power generated using renewable energy sources as the world races to reduce the carbon footprint of the internet [5]. Green energy sources are however not solely reliable as they are weather- and location-dependent. To output, an end-to-end efficient utilization of the harvested renewable energy that can sustain the high demands of users for low latencies and communication reliability, mobile edge computing (MEC) has been proposed. It consists of the implementation of computing and storage capabilities at the access edge of the network. Several benefits are realized through the implementation of the MEC approach. Aside from the assurance of low latency communications, the backhaul traffic load is drastically reduced since data is fetched from edge servers. Also, the quality of the immersive media content can be adapted to the quality of a user's channel [6], [7].

The work in [8] studied the numerous energy-efficient compromises for green communication and shows the impact of these compromises on multiple types of wireless network architectures. Various energy-efficient schemes that can improve the architecture were also studied. Efficient energy utilization in green wireless access infrastructure is surveyed to exploit the base station sleep mode approach in [9]. In [10], greener and sustainable networks aimed at achieving energy sufficiency were studied. The authors explored the energy consumption of user devices and presented a review of improvement approaches. The study was not extended to cover renewable energy and cellular networks in detail.

Despite the current efforts, the explosive growth of IoT devices and the use of the access network as a communication enabler also pose new challenges for energy-efficient communication. Works in [5], [11] investigated the subjective expected experience of 5G users in mobile augmented reality applications. This was done through the construction of 5G user experience models for mobile augmented reality and tactile internet, as well as mobile edge cloud computing powered by renewable energy to model the power consumption, minimization model.

The importance of energy harvesting in the plan for sustainable mobile networks and appropriate energy management techniques are explored in this paper to efficiently utilize the constrained energy to achieve maximum cache hit ratio and low latency in content delivery. We propose an energy-saving caching and fetching model to decide the optimal cache placement and fetching in the likeness of a multi-constrained multiple-knapsack optimization problem. The specific contribution of this paper to literature are summarized as follows:

- We present an energy consumption model for the caching and transmission of immersive videos whereby different parts of the same content file are cached among a cluster of edge servers.
- We model the content placement problem of the clustered constrained edge servers as a multi-constrained multiple-knapsack problem to maximize the cache hit rate under minimum energy expenditure and low latency constraints for caching and transmitting content.
- We propose a cache-maximization energy-efficient heuristic algorithm to find the optimal cache placement solution.

The rest of the paper is arranged as follows: Section II describes the system model. In Section III, the problem formulation and steps towards a feasible solution are presented. Section IV discusses the results and performance evaluation of the proposed scheme. The paper concludes and highlights future work in Section V.

II. System Model and Problem Formulation

In this section, the network model and key definitions used in this paper are presented. The system architecture consists of a regionally organized cloud (ROC), MEC access nodes (MANs), renewable energy sources, and User Equipment (UE) as depicted in Figure 1.

A. System Model

We consider a geographical location having several MANs providing network access to users with similar preferences and group them in a cluster under a ROC. ROCs have sufficient computing and storage resources, therefore most of the media contents are kept on the ROC servers unless requested by a user. A ROC has a set of clustered base stations each with edge computing and storage capabilities, $M_s = \{m_1, ..., m_S\}$. The users $U = \{u_1, ..., u_P\}$ are randomly placed in the access network with the expectation of receiving the best service.

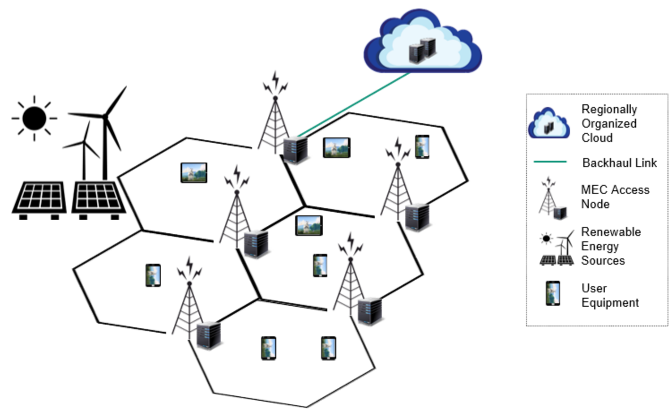

Fig. 1. Renewable Energy Powered Edge Caching Network

The set $F = \{f_1, f_2, ..., f_V\}$ is the content files to be cached where the probability of a file being requested is expressed by its popularity, following the Zipfian distribution. The cache size of a MAN server is given by $C_{\max}$, where the utilized cache space is defined as $\vartheta_{s\in S} \times s_{f_v}$ where $\vartheta_{s\in S}$ denotes the total number of contents cached and s_{f_v} is the content size. The maximum data transfer rate is $R_{\max}$. A request for a file ρ_{f_v} can be served directly by the local MAN of the user, a neighbor MAN server, or the ROC [12]. The energy consumption in the edge caching network is proportional to the energy dissipated on content caching, transcoding, and transmission.

1) Content caching energy

The energy cost of caching content f_v on the MAN server s is given by $\varepsilon^c_{f_v} = s_{f_v} \times H_\varepsilon$, where H_ε is the required hardware energy in $watt/bits$. A binary cache decision variable is defined as $c^{f_v}_{m_s} = \{0,1\}$. $c^{f_v}_{m_s} = 1$ is the cache decision that f_v is on the MAN server m_s , and $c^{f_v}_{m_s} = 0$ is the state that the file f_v is not cached on m_s . Hence, the required caching energy is $c_{\varepsilon s} = \varphi^{f_v}_{m_s} \varepsilon^c_{f_v}$. Similarly, $c^{f_v}_{m_k} = 1$ is the state that f_v is cached on neighbor MAN server m_k . $c_{\varepsilon k} = \varphi^{f_v}_{m_k} \varepsilon^c_{f_v}$ is the caching energy at server m_k . Total caching energy is $c_T = c_{\varepsilon s} + c_{\varepsilon k}$

2) Content transmission energy

The transmission energy required to fetch a user's request is dependent on where the content can be found. $\varphi_{m_s}^{f_v} \in \{0,1\}$ is the transmission decision variable that f_v is transmitted from the MAN server m_s. If the requested content is served by a neighbor MAN server, then $\varphi_{m_k}^{f_v} = 1$ represents the MAN server k serving the content. If $\varphi_{m_s}^{f_v} = 1$, the data rate for transmission is defined as $R_{f_v}^{s} = \log_2(1 + SINR(p,s,f))$, where $SINR(p,s,f)$ describes the signal to interference noise ratio. On the other hand, if the content is transmitted by a neighbor MAN server, then its transmission rate is $R_{f_v}^{k} = \log_2(1 + SINR(p,k,f))$. Upon delivery of the requested file, the total throughput is computed as:

$$\mathrm{T} = (\varphi_{m_s}^{f_v} \times R_{f_v}^{s}) + [(1 - \varphi_{m_s}^{f_v}) \times R_{f_v}^{k}] \tag{1}$$

The cache hit ratio of the system represents the number of video requests served by any MAN server. Its given by $H_C = \varphi_{m_s}^{f_v} + c_{m_s}^{f_v} + \sum_{k \neq s} \varphi_{m_k}^{f_v} + c_{m_k}^{f_v}$.

Finally, the total edge caching network energy consumption can be calculated as:

$$\Upsilon = \sum_{s \in S} \sum_{f \in F} \mathrm{T} + c_T \tag{2}$$

B. Problem Formulation

Immersive media occupies larger storage sizes because they are multidimensional, and thus require high bandwidth data transfer rates for their transmission. Our primary objective is to increase the energy efficiency of the renewable-source-powered edge caching network by minimizing the total consumed energy as described below. Equation (3) is the global objective function that aims to minimize the overall energy consumption of the network. The Constraint in (4) bounds the energy required to fetch content f_v by the maximum defined transmission energy R_{max}. The data rate for transmission of f_v is greater than the threshold R_{TH} as depicted as a constraint in (5). The constraint in (6) specifies that the cache must not go beyond the allotted maximum storage, and (7) restricts caching a file on M number of cache servers. The last set of constraints depicted by (8) and (9) restrict the caching decision and cache reachability to non-negative values.

$$\text{Maximize} \quad \sum_{s \in S} \sum_{f \in F} \mathrm{T} + c_{\varepsilon} \tag{3}$$

$$\text{Subject to:} \quad \sum_{p \in M_s} \varphi_{m_s}^{f_v} \varepsilon_{f_v}^{c} \leq R_{\max} \qquad \forall f \in F \tag{4}$$

$$\sum_{s \in S} \varphi_{m_s}^{f_v} R_{f_v}^{s} \geq R_{TH} \qquad \forall s \in S \tag{5}$$

$$\sum_{f \in F} c_{m_s}^{f_v} s_{f_v} H_{\varepsilon} \leq C_{\max} \qquad \forall s \in S \tag{6}$$

$$\sum_{e=1}^{|ET|} \varphi_{m_s}^{f_v} \leq M \qquad \forall s \in S \tag{7}$$

$$c_{m_s}^{f_v} \in \{0,1\} \qquad \forall s \in S \tag{8}$$

$$\varphi_{m_s}^{f_v} \in \{0,1\} \qquad \forall s \in S \quad \forall f \in F \tag{9}$$

III. ENERGY-AWARE PART-BASED EDGE CACHING SCHEME (EAPBEC)

The energy-aware edge caching problem is presented as a multi-constrained multiple-knapsack problem, where the cluster of MAN edge servers are the multiple knapsacks, the profit of caching a file is the size of the file, and the weight is the energy consumed for caching and transmitting the files. An optimal solution to a large-scale energy minimization is difficult to find in a feasible time. Algorithm 1 presents the file splitting process and its placement at the edge cache. First, previous request patterns and popularity counters of each file are checked. Then, caches of the MAN servers are initialized. The ROC prepares the list of the V most popular files to be cached. The energy consumed for caching a file is calculated to update the file catalog for future cache budgeting.

```
Algorithm 1        File Splitting and Placement Algorithm
Input: New video f_v from the ROC server.
Initialize: Caches on MAN servers M_s = {m_1,...,m_S} .
    while ε^c_{f_v} = s_{f_v} × H_ε do
        for each file f_v do
            if s_{f_v} ≤ SF then
                No file splitting needed: Cache f_v on
                MAN server s .
            else if SF ≤ s_{f_v} ≤ MF then
                Split f_v into two parts: f_v^1 and f_v^2 Cache
                only f_v^1 in quality q at MAN server s .
            else if MF ≤ s_{f_v} ≤ LF then
                Split f_v into h parts for each h ≥ 500 MB
                Cache only f_v^1 at MAN server s and f_v^2 at
                MAN server k .
            end if
        end for
    end while
```

The transmission energy required for transmitting each file is also calculated. Since immersive media files are large, the ROC splits the files into parts (small file $SF \leq 500$ MB; medium file $500\text{ MB} < MF \leq 1\text{ GB}$; large file $LF > 1\text{ GB}$), then distributes the parts among the neighbor servers. In this way, a large size file does not occupy much cache space of the edge server. Serving a file request is mostly done locally. If the part of a file is not in the cache, a *get file* request is initiated to all neighbor MANs for file serving assistance. A user is then served if the desired part of the file is cached among the cluster of MANs. If the file is not in the cluster cache, the ROC is requested to serve the file. The parts of a requested file are served in sequential order when and only when the user responds to *continue to play* notifications. The moment there is no response from a user, the streaming session is ended to save the energy and bandwidth of the network.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
No. of MAN servers in a cluster	4
Cell coverage size	200 m
MAN server cache capacity	10 GB
Total number of video files	10000
Transmission power	[40 ~ 47] dBm
Channel bandwidth	20 MHz
Noise power	-95 dBm
Delay between MAN servers	20 ms
Delay between ROC and MAN servers	200 ms
Zipf's parameter	0.45
Average request rate	[0.2 ~ 1]

Algorithm 2 File Fetching Algorithm

Input: File request f_v on MAN server s by a user u_p.

Output: Cache Hit Ratio.

Initialize: Available $R_{f_v}^s$.

Initialize: Cache on MAN servers $M_s = \{m_1, ..., m_S\}$.

for $n \in 1,...,N$ do
  for each request r_{u_k} on MAN s do
    if $\varphi_{m_s}^{f_v} == 1$ then
      Serve u_k from MAN server s.
      $R_{f_v}^s = \log_2(1 + SINR(p, s, f))$
    else if $\varphi_{m_k}^{f_v} == 1$; $k \neq s$ then
      Fetch f_v from MAN server k and serve u_k.
      $R_{f_v}^k = \log_2(1 + SINR(p, k, f))$
    end if
  switch (Check s_{f_v})
    case $s_{f_v} \le SF$
      Complete fetching session
    case $SF < s_{f_v} \le MF$
      Send f_v^2
    case $MF < s_{f_v} \le LF$
      Fetch remaining parts of f_v and deliver to MAN server s.
    end case
  end for
end for

IV. PERFORMANCE EVALUATION

The simulations are per the implementation of an Edge Simulator considering many adjustable network and content parameters [13]. Table I presents the specified parameters used in the experiments to evaluate the proposed scheme. In the first set of experiments, we find the percentage of the cache hit ratios under varying network power consumption. We use limited caching and transmission power at average peak times to determine the performance of the proposed scheme as compared with the least frequently used (LFU), first-in-first-out (FIFO), and least frequently used (LRU). The proposed EAPBEC outperforms the LFU, FIFO, and LRU with performance gains of 20%, 46%, and 49% respectfully, as shown in Figure 2. The EAPBEC scheme shows a steady improvement with the varying consumption power.

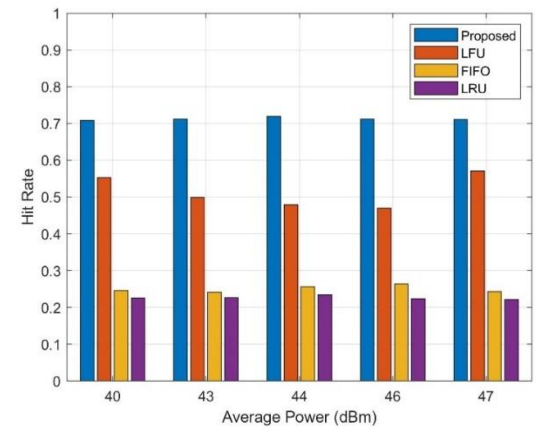

Fig. 2. Cache hit ratio

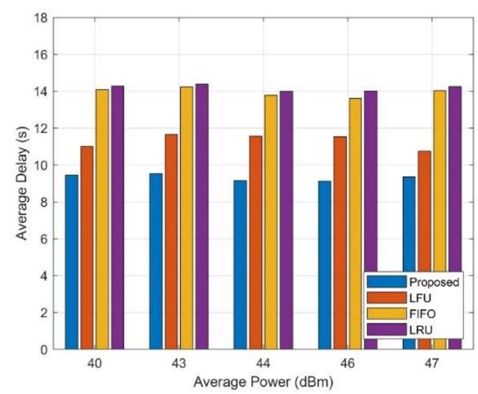

Fig. 3. Average delay

The second set of experiments demonstrated in Figure 3 shows that the proposed scheme achieves the lowest delay for content fetching and delivery. The performance is quite steady with the varying transmission rates and energy. The average network delay is lowest for the proposed scheme because it efficiently caches most of the requested parts of the content files among the MAN servers. The LFU, FIFO, and LRU performs no file splitting but caches wholesome video files so the cache is filled with only a few contents. A cache miss registers assistance from the ROC to serve the request, with the non-negligeble delay between the ROC and the MAN servers undesirably affecting the network responsiveness.

Figure 4 presents the performance graph of the backhaul traffic versus varying request rates of the schemes. The backhaul traffic generated when requests are served using the proposed EAPBEC scheme is incommensurate with the projected performance. The FIFO scheme has the best performance in this comparison, with the proposed scheme just ahead of LFU. This can be as a result of the splitting of larger files leaving several parts of the file on the ROC for transfers when users continue to engage the media. If all parts of a file are not cached among the clustered MAN servers, the remaining parts are stored on the ROC for future requests which incur additional data traffic on the backhaul network during delivery. More results and discussions from the simulations in this research will be presented in the extended version of this paper.

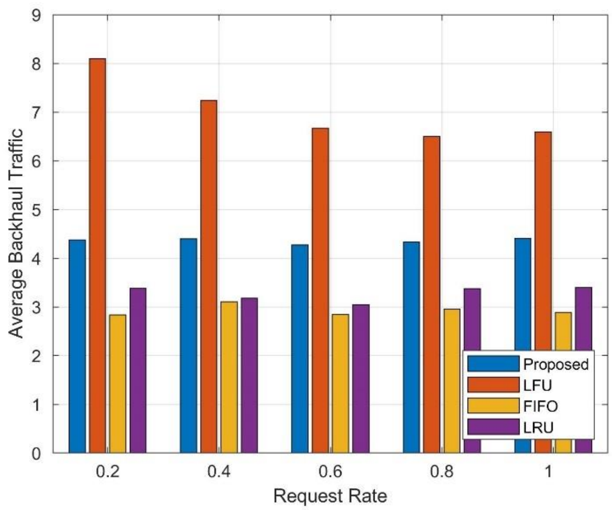

Fig. 4. Average backhaul traffic

V. Conclusion and Future Work

The information-communications technologies (ICT) industry contributes up to 3.9 % of all global carbon emissions. The growth of the ICT industry has been one of the most progressive technological and societal developments of the last decade, with new bandwidth-critical applications and services projected to drive this growth further and faster. In this work, we presented an energy consumption model for the caching and transmission of immersive videos whereby different parts of a content file are cached among a cluster of edge servers. The proposed scheme achieved high cache hit ratios and low delays in comparison with the LFU, FIFO, and LRU.

Applications and services of the immersive and interactive media (VR, AR, 360° video, Multiview video) continues to grow at a faster pace with the advancement of devices capable of delivering all kinds of content presentations. In our future work, we will study new means of optimizing the resource allocation for green energy-aware caching networks for seamless immersive media streaming and caching.

References

[1] CISCO, “2020 Global Trends Report,” *Cisco*, pp. 1–95, 2020.

[2] R. Carter, “What is Immersive Media: An Introduction,” 2021. https://www.xrtoday.com/mixed-reality/what-is-immersive-media-an-introduction/ (accessed Feb. 10, 2022).

[3] S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, “Edge Caching and Computing in 5G for Mobile AR/VR and Tactile Internet,” *IEEE Multimed.*, vol. 26, no. 1, pp. 21–30, 2019, doi: 10.1109/MMUL.2018.2879591.

[4] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. Blair, and A. Friday, “The climate impact of ICT: A review of estimates, trends and regulations,” no. December, 2021, [Online]. Available: http://arxiv.org/abs/2102.02622.

[5] Y. Cheng, “Edge caching and computing in 5G for mobile augmented reality and haptic internet,” *Comput. Commun.*, vol. 158, pp. 24–31, 2020, doi: 10.1016/j.comcom.2020.04.054.

[6] F. Song *et* al., “Smart collaborative caching for information-centric IoT in fog computing,” *Sensors (Switzerland)*, vol. 17, no. 11, pp. 1–15, 2017, doi: 10.3390/s17112512.

[7] C. Sonmez, A. Ozgovde, and C. Ersoy, “Performance evaluation of single-tier and two-tier cloudlet assisted applications,” *2017 IEEE Int. Conf. Commun. Work. ICC Work. 2017*, pp. 302–307, 2017, doi: 10.1109/ICCW.2017.7962674.

[8] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. Ul Hassan, and C. Yuen, “Energy efficiency tradeoff mechanism towards wireless green communication: A survey,” *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1. 2016, doi: 10.1109/COMST.2015.2490540.

[9] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, “Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey,” *IEEE Commun. Surv. Tutorials*, vol. 17, no. 2, pp. 803–826, 2015, doi: 10.1109/COMST.2015.2403395.

[10] J. Lorincz, A. Capone, and J. Wu, “Greener, Energy-Efficient and Sustainable Networks: State-Of-The-Art and New Trends,” *Sensors*, vol. 19, no. 22, 2019, doi: 10.3390/s19224864.

[11] J. Liao, K. K. Wong, Y. Zhang, Z. Zheng, and K. Yang, “Coding, Multicast, and Cooperation for Cache-Enabled Heterogeneous Small Cell Networks,” *IEEE Trans. Wirel. Commun.*, 2017, doi: 10.1109/TWC.2017.2731967.

[12] O. Serhane *et al.*, “Energy-aware Cache Placement Scheme for IoT-based ICN Networks To cite this version : HAL Id : hal-03161227 Energy-aware Cache Placement Scheme for IoT-based ICN Networks,” 2021.

[13] C. Wang, R. Li, W. Li, C. Qiu, and X. Wang, “SimEdgeIntel: A open-source simulation platform for resource management in edge intelligence,” *J. Syst. Archit.*, vol. 115, p. 102016, May 2021, doi: 10.1016/j.sysarc.2021.102016.